Data Analysis with SQL

Project Deliverable

- Your deliverable will be a notebook with your solution.

Instructions

Over the past few years, ride-sharing apps have been on the rise across many cities in the world. While this has happened, Uber and Lyft's ride prices are not constant like public transport. They are greatly affected by the demand and supply of rides at a given time.

As a Data Scientist working to understand this market, you have been tasked to come up with a descriptive analysis report to help a Ride-Sharing Startup coming into this space, understand the various patterns on how pricing works for the existing ride-sharing company.

Luckily, you were able to access some real-time data from Uber & Lyft's API and weather data from Weather API conditions.

You build a custom application in Scala to query data at regular intervals and saved it to DynamoDB. The queried cab ride estimates are done after every 5 mins and weather data after every 1 hr.

The cab ride data covers various types of cabs for Uber & Lyft and their price for the given location. Weather data contains weather attributes like temperature, rain, cloud, etc for all the locations taken into consideration.

Now that you have your data in the given dataset, write SQL queries to perform descriptive analysis highlighting key insights that would be helpful in helping the startup develop a new product.

Hint:

Your first guess on this would be to think about the time of the day; whether times around 9 am and 5 pm should see the highest surges on account of people commuting to work/home.

Dataset

- Dataset Download URL = https://bit.ly/3dZiVp8

Glossary

Cab Rides Dataset

- distance: the distance between source and destination.
- cab_type: Uber or Lyft
- time_stamp: epoch time when data was queried
- destination: destination of the ride
- source: the starting point of the ride
- price: price estimate for the ride in USD
- surge_multiplier: the multiplier by which price was increased, default 1
- unique identifier
- product_id: uber/ lyft identifier for cab-type
- name: Visible type of the cab eg: Uber Pool, UberXL

Weather Dataset

- temp: Temperature
- location: Location name
- clouds: Clouds
- pressure: pressure in mb
- rain: rain in inches for the last hr
- time_stamp: epoch time when row data was collected
- humidity: humidity in %
- wind: wind speed in mph

Project Source: https://bit.ly/2AKnlBL